Changing patterns of self-citation: Cumulative inquiry or self-promotion?

Ken Hyland and Feng (Kevin) Jiang

University of East Anglia, UK & Jilin University, China

Corresponding author:

Prof Ken Hyland
School of Education and Lifelong Learning
University of East Anglia
Norwich, UK

Email K.hyland@uea.ac.uk
Phone 01227 458972

Keywords self-citation; academic discourse; diachronic change; disciplinary differences; assessment culture; higher education

Running head: Changing patterns of self-citation

Counts (excluding abstract, bios & title page but including references and appendix)
characters (no spaces) 45,428 **Words** 8085

Author Biographies

Ken Hyland is Professor of Applied Linguistics in education at the University of East Anglia. He was previously a professor at University College London and the University of Hong Kong and has taught in Africa, Asia and Europe. He is best known for his research into writing and academic discourse, having published over 200 articles and 26 books on these topics with 35,000 citations on Google Scholar. Address: School of Education, UEA, NR4 7TJ, UK. Email: k.hyland@uea.ac.uk

Feng (Kevin) Jiang is Distinguished Professor at the Faculty of Foreign Language Education at Jilin University, China. He gained his PhD under the supervision of Professor Ken Hyland at the Centre for Applied English Studies at the University of Hong Kong. His research interests include disciplinary discourse, corpus studies and academic writing and his publications have appeared in most major applied linguistics journals. Address: #2699 Qianjin Street, Changchun, Jilin, China. Email: kevinjiangfeng@gmail.com

Changing patterns of self-citation: Cumulative inquiry or self-promotion?

Abstract

Self-citations are a familiar, if sometimes controversial, element of academic knowledge construction and reputation-building, contributing to both the cumulative nature of academic research and helping writers to promote their scientific authority and enhance their careers. As scholarly publications become more specialised, more collaborative and more important for promotion and tenure, we might expect self-citation to play a more visible role in published research and this paper explores this possibility. Here we trace patterns of self-citation in papers from the same five journals in four disciplines at three-time periods over the past 50 years, selected according to their impact ranking in 2015. We identify a large increase in self-citations although this is subject to disciplinary variation and tempered by a huge rise in citations overall, so that self-citation has fallen as a proportion of all citations. We attempt to account for these changes and give a rhetorical explanation for authorial practices.

Changing patterns of self-citation: Cumulative inquiry or self-promotion?

1 Introduction

Self-citations, where the citing and cited paper have at least one author in common, account for a substantial share of all academic citations. The cumulative nature of scientific research perhaps makes this kind of embedding a natural part of the communication process as writers' earlier work informs their current research (e.g. Costas et al. 2010). Equally, however, the practice is often condemned as vanity and self-promotion, claiming an author's scientific authority or gaming a system which rewards citation. Because of this, there have been calls to exclude them from the statistics which measure the impact of academic literature and which, in turn, contribute to tenure, promotion and funding decisions (e.g. Fowler & Aksnes 2007). Despite this controversy, the citation indexes generally include self-citations and, as research becomes ever more specialised, co-authorship more common and competition to get published more intense, self-citation may now be a more significant rhetorical and tactical tool in the struggle for visibility and authority.

In this paper we examine this idea, tracing patterns of self-citation and seeking to account for the changes that have occurred. First, we discuss the characteristics and importance of self-citation and how it has grown in recent years. We then outline our corpus and approach before going on to explore changes in the use of self-citation across four different disciplines over the past 50 years, attempting to give a socio-rhetorical explanation for these changing authorial practices.

2 Self-citation: practices, perceptions and problems

The use of self-citation is the principal means by which writers can emphasise the linear progress of their research to demonstrate consistent lines of inquiry and the development of their ideas. This is what Cronin (1981:16) long ago described as "frozen footprints in the landscape of scholarly achievement, footprints which bear witness to the passage of ideas". It also, however, helps authors to rhetorically construct professional credibility, establish authority in an area and generate interest in their earlier research. These diverse functions, together with the growing importance of bibliometric measures of impact based on citation counts, mean that the topic has generated considerable research interest.

While applied linguists have analysed citation to show differences in disciplinary rhetoric (e.g. Hyland 2003; Harwood 2009) and sociologists have offered insights into the social construction of knowledge (e.g. Gilbert & Mulkay 1984; Bornmann & Daniel 2008), citation practices have mainly been studied by those working in bibliometrics. Prichard (1969) states that bibliometrics is "the application of mathematical and statistical methods to books and other media of communication" (p. 349). Thus, in this discipline researchers conduct statistical analyses of the academic literature to compile citation indexes and determine the popularity and impact of specific articles, authors, and publications. This concern with the relationships between texts, and between texts and writer productivity, emphasizes that scientific papers are embedded in a literature and that writers are linked into wider social networks. It thus draws attention to the role that publication plays as the institutionalized system which both creates knowledge and distributes rewards, underlining the importance of reputation to academic endeavour.

The need to have one's work recognized and cited by others is, in fact, an increasingly valued commodity in today's fiercely competitive academic world. Rivals seek to better each other in the quality and significance of their work and in the esteem in which they are held by their professional colleagues because this is a key factor in further career moves. Bibliometric evaluation exercises themselves, such as the Research Excellence Framework in the UK, which emphasize physical production over scholarly reflection, neatly close this circle of publish-citation-reward.

Given the high-stakes nature of citation counts as measures of productivity and expertise, users of bibliometric results sometimes condemn self-citations as a means of artificially inflating citation rates and thus strengthening the authors' position in the scientific community. While some bibliometric studies of large scale data suggest that regularities of self-citation are a natural part of scientific communication (e.g. Bonzi & Snyder 1991; Harzing 2013), analyses at the levels of individuals, research groups and institutions indicate that self-citation distorts citation counts and reduces their reliability as a measure of quality (Thijs & Glanzel 2006). In a study of more than half a million citations to articles by Norwegian authors in the Science Citation Index, for instance, Fowler & Aksnes (2007) found that

the more one cites oneself the more one is cited by others, with each self-citation generating an additional three citations (though not necessarily to the same paper) from other scholars over a five-year period.

Studies consistently show that self-citations account for between 10% and 35% of all citations, depending on the field and the stage of development of the area (e.g. Rousseau 1999; King, Correll, Jacquet & Bergstrom 2015). Using a three-year citation window, for instance, Aksnes (2003) found 36% of the citations in 45,000 publications by Norwegian authors represented self-citations and Glanzel et al. (2006) note that 27.2% of all citations received by articles and notes recorded in the Science Citation Index were self-citations in the citation window 2001-2003. Peritz and Bar-Ilan (2001) discovered that five papers published in *Scientometrics* in 1990 and another five in 2000 contained over 50% of self-citations and noted that about 30% of the papers in both years had rates of 20% or more. At the extreme end of the scale, one paper in the journal *Science* entitled *A Comparison of Whole-Genome Shotgun-Derived Mouse Chromosome 16 and the Human Genome* includes 220 self-citations to previous papers by the 175 authors. There are, however, few articles which do not include self-citations and studies show that authors tend to cite their own work more heavily than that of others (White 2001).

Self-citation is shown to be particularly heavy among authors who have a long history of engagement in an area (Pichappan & Sarasvady 2002) and this increases as scholars move through their careers and publish more research (Chang 2006). It is also more frequent in the work of those who are seeking to publish outside their specialist field (Cronin & Shaw 2002) and comprises a higher share of citations in papers that are themselves poorly cited (Aksnes 2003). Self-citation also rises with the number of authors, although multiple authorship mainly increases the probability of being cited by others (Glanzel & Thijs 2004). In addition, men seem to cite their own papers at a higher rate than women, by 56% from 1779 to 1991 and by 70% since then (King et al. 2015). Because citation follows a pattern of preferential attachment, with new citations referring to papers that are already popular (Maliniak et al. 2013), this may also create gender differences in the number of citations received from other authors. Most strikingly, self-citation is most common among authors who are themselves heavily

cited. Phelan (1999) discovered that only two of the 56 most highly cited authors in education did not cite themselves over a twelve-year period, while 154 of the 280 citations (55%) received by one author were the result of self-citation.

Finally, self-citations, like all citations, are characterized by large inter-disciplinary variations (Hyland 2003; Aksnes 2003; King et al. 2015). Aksnes (2003), for instance, found the lowest percentage in Clinical medicine (17%) and the highest in Chemistry and Astrophysics (31%) while the Public Policy Group (2011) reports figures ranging from a high of 42% for engineering, through 32% for language and communication to a low of 21% for medical and life sciences. The nature of an academic subject, its social and rhetorical practices, the kinds of publications it favours, and the characteristics of its argument patterns all distinguish academic disciplines and complicate comparisons. Generally, the social sciences and humanities have low rates with a fifth to a quarter of citations being self-cites, whereas in the STEM disciplines (Science, Technology, Engineering and Mathematics) the rate is around a third (Hyland 2003; Snyder & Bonzi 1998). The extent of the variation is likely to be determined by the serial nature of this work with considerable specialization in engineering and the sciences with problems pursued very intensively, often by one or two labs and with relatively few competitors. In such contexts self-citation is inevitable and facilitates academic development. The humanities, on the other hand, contains disciplines where cumulative work by a single team or lab developing distinctive ideas and approaches is far less important.

Clearly, motives for self-citation vary (Hyland 2003; Safer & Tang 2009) with Hyland's respondents conveying a strong desire to establish their authority in their field and display lines of priority in research. In general, some research has indicated that there are no significant differences in the motivations for citing between self-citations and citations to other works (Bonzi & Snyder 1991). Overall, the use of self-citations is a major way in which writers can seek to rhetorically underpin their professional reputations, emphasizing their contributions and connectedness. This does not mean, however, that self-citation indicates manipulative behaviour. The *absence* of self-citation might simply suggest a lack of influence or consistency of research direction (Glanzel et al. 2006) or that the author is a junior scholar with little prior work to cite (Harzing 2013).

Self-citation may therefore be closely related to the relentless increase in specialisation in research, which Ruscio (1985) accounts for thus:

> There are epistemological reasons: the sheer volume of knowledge and its rapid expansion compel a scientist to carve out his own niche of expertise. There are also sociological reasons: academics achieve status within the profession by advancing knowledge, a dynamic that requires precise contributions. Institutions of higher education themselves compete for status, reinforcing the individual's motivation. (cited in Becher & Trowler 2001: 66).

Authors may therefore have good reasons to cite their own work as it can demonstrate programmatic research over an extended period with individual studies building upon each other and referring back to their earlier work. The fact that older researchers do more self-citing supports this 'serial development of knowledge' perspective (Public Policy Group 2011), although individual practices suggest that self-citations do not necessarily reflect the importance of the research it reports or its impact beyond the interests of its own producers (Schubert et al. 2006).

Self-citation, then, is of interest to researchers in a number of fields and for numerous reasons. Its importance in contributing to academic argumentation, knowledge-building and scholarly career trajectories make it a fascinating subject for applied linguists and writing researchers. The additional weight given to citations in institutional evaluations in the last 20 years only increases the significance of self-citations and for this reason alone we might expect it to have increased in recent times, although these increases are likely to vary by academic field. In the following sections we explore changes in the use of self-citation across four different disciplines over the past 50 years. First, we outline our corpus and approach.

3 Corpora and procedures

Bibliometric tools for analysing citations such as *CiteSpace* and *VantagePoint* automatically harvest citations and subject them to mathematical models and statistical analysis. Because we are not seeking to trace author networks, map collaborative research or quantify individual citing behaviour, our methods are more modest. In addition, bibliometric software often seeks to track particular individual writers or teams and so only counts articles that include the citing-cited pairs. Equally importantly, it

generally only performs its analyses on citing articles, not on actual citations and as a single article often contains more than one citation to an academic's work, this is likely to skew the instances which are reported. Our procedure was to construct three corpora using very specific criteria and count actual instances of self-citation manually using a concordancer.

We created the three corpora taking research articles from the same five journals in four disciplines spaced at three periods over the past 50 years: 1965, 1985 and 2015. The different time spans were chosen to see if any changes were more pronounced in the later or earlier period, although we were concerned with overall changes over the 50 years. Applied linguistics, sociology, electrical engineering and biology were selected as representative of the social sciences, humanities and hard sciences, and we took six papers (three single authored and three co-authored) at random from each of the five journals which had achieved the top ranking in their field according to the 5 year impact factor in 2015.[1] The journals are listed in Appendix 1 and together the corpus comprised 360 papers of 2.2 million words as shown in Table 1

Table 1: Corpus size and composition

Discipline	**1965**	**1985**	**2015**	**Overall**
Applied linguistics	110,832	144,859	237,452	493,143
Biology	244,706	263,465	237,998	746,169
Electrical engineering	92,062	97,545,	235,681	425,288
Sociology	149,788	196,232	262,203	608,223
Totals	597,388	604,556	973,334	2,272,823

The table indicates the huge increase in the length of articles over this period, with the number of words rising by 63%. Only figures for biology have remained steady, opening the period with far longer papers than our comparison disciplines when the field was enjoying phenomenal success with research on the genetic code, the structure of hemoglobin and myoglobin and work on the cell. The length of papers in Electrical engineering, in contrast, has exploded since 1965, perhaps because of the massive changes in the discipline resulting from the invention of transistors (Rowe 2016).

[1] Two journals, *TESOL Quarterly* and *Foreign Language Annals* only began in 1967 and so papers were chosen from issues in that year. *College Composition and Communication* is the only one not ssci-indexed but it is an influential, long running journal listed in the language & linguistics sub-category of Arts & Humanities Thomson Reuters citation index.

Each of the three time corpora were then searched individually for citational forms such as a name or date in brackets (1), a number in squared brackets (2), superscript references (3) and Latinate references to other citations (4) using the concordance software *AntConc* (Anthony 2011).[2] Every reference to one of the authors' earlier works was counted as a self-citation, with two or more texts in a single mention tallied separately.

(1) In areas of applied linguistics such as language learning and teaching, there is, as I noted elsewhere (Block, 2012, 2014), a dearth of class-based research. (Applied linguistics)

(2) As we demonstrate in [21], for multimodal data fusion, the use of a flexible algorithm like EBM better optimizes the ICA criterion… (Electrical engineering)

(3) Such a model attributes differences in mobility solely to differences in the marginal distributions of the tables, which, following Hope[30] we may identify as reflecting the distribution of origins and of available occupations. [Sociology]

(4) humans have a highly developed ability to learn new behaviors from observation, but we are not automatically successful, and true behavioral mimicry is exceedingly difficult (ibid). [Sociology]

Next, we distinguished whether authors referred to themselves explicitly in the reporting clause or whether they simply noted an earlier paper in parenthesis without such self-mention. The difference between *sentential reference* (5) and *parenthetical reference* (6)[3] is illustrated here (the authors of the article are underlined):

(5) In an earlier study (Jeffery & Wilcox, 2014), we focused on the perspectives of native English-speaking students using an analytic framework.

(6) Recent investigations (e.g. Fitzpatrick 2007; Zareva and Wolter 2012) have questioned the validity of assuming there is a coherent norm behaviour in native speakers.

[2] Authors self-citing are underlined in examples.
[3] Swales (1990) refers to these as *integral* and *non-integral* citations.

Following Ioannidis (2015), we further distinguished between *direct* and *co-author* self-citation.[4] Direct self-citation is the classic example of an author citing his or her earlier papers, those self-citations that one researcher receives from him/herself. Co-author self-citation refers to situations where the co-authors cite a paper which not all the current authors wrote. Thus the first kind occurs whenever A is also co-author of a paper citing a publication by A while in the second type A may not be in the by-line of the citing paper. So, in Wilson and Brommer's paper (7), the citation to a previous paper they have co-authored with Nussey is a co-author self-citation:

(7) Increasingly, within-individual variation is being integrated into evolutionary theory (e.g. Nussey, Wilson & Brommer, 2007), but major gaps exist in our knowledge of processes affecting this level in the hierarchy of variance.

Nussey has not directly self-cited himself/herself here, but Wilson and Brommer have and this can take on substantial proportions when there are many co-authors. In other words, while these are direct self-citations for some of the co-authors, not all authors have directly self-cited. The distinction is used by bibliometricians to track the behaviour of particular writers and authorial networks, and while it makes little difference to the overall counts in this study, we find it a useful distinction for discussing self-citing behaviour overall. We discuss the results in the following sections.

4 Self-citation: overall results

We found over 1,450 self-citations in the 2015 corpus, representing 15 cases per 10,000 words of text and comprising about 11% of all citations. Table 2 shows a steady and substantial increase since 1965, with raw figures increasing by 137% and by 45.6% when adjusted for the large rise in the length of papers. Clearly, self-citation remains a key feature of academic writing and, as discussed in the next section, exhibits considerable disciplinary differences.

[4] Ioannidis actually identifies other types including collaborative, where only a subset of a larger team cited as authors in a large number of papers, and coerced, where reviewers ask for their papers to be cited in a submission to a journal. Neither is available to us for study.

Table 2: Self-citations over time

	1965	1985	2015	50 year change
Self-citations	618	903	1464	136.9%
Self-citations per article	5.2	7.5	12.2	134.6%
per 10,000 words	10.3	14.9	15.0	45.6%
Self cites/all cites	15.2%	15.2%	10.9%	-28.0%

The rise in raw numbers has occurred mainly in the last 30 years with the expansion of academic publishing, but the real surge was in the earlier period where the growth of self-citation significantly outstripped the increase in words per article. This figure has, however, remained steady at around 15 per 10,000 per words since then. Despite the overall rise, self-citation has declined as a proportion of citations overall, down 28% during the period, as both citations, and the literature to cite, have grown enormously.

Thus the rise in self-citation is not keeping pace with citations in general, which have more than doubled in these four corpora over the last 50 years (Hyland & Jiang 2016). There are more scholars publishing now than ever before, with more journals, more articles and more publishers involved in meeting, and generating, the demand to publish research. Thus Thompson Reuters (2012) report a four-fold increase in submissions to the 4,200 journals using the *ScholarOne* manuscript processing system between 2005 and 2010 and some estimates suggest that about one million authors publish some 1.5 million articles each year in 25,000 peer-reviewed scholarly journals (Ware & Mabe 2012). Presumably, then, the huge volume of research now available, and more easily accessed through online publication, provides authors with a greater number and variety of sources which can, and need to, be cited.

We also found significant changes in the ways authors chose to embed their references to earlier work, with a shift to parenthetical references (with no mention of the authors in the accompanying sentence) and to co-author self-citation (citations that one author receives from his/her co-authors in the cited document). Table 3 presents the changes as both raw frequencies and cases per 10,000 words.

Table 3: Types of self-citation over time

	1965		1985		2015		Changes	%
Reference type	**No.**	**per 10,000**	**No.**	**per 10,000**	**No.**	**per 10,000**	**No.**	**per 10,000**
sentential	42	0.7	27	0.4	41	0.4	-2.4	-42.9
parenthetical	576	9.6	929	15.4	1353	13.9	134.9	44.8
Direct only	382	6.4	390	6.5	379	3.9	-0.8	-39.1
Co-author	236	4.0	513	8.5	1085	11.1	359.7	182.2

The large increase in parenthetical self-citation is surprising as we might expect that pressures on academics to promote themselves for career purposes would lead to more self-mention rather than less, although raw frequencies have increased slightly in recent years. However, a parenthetical self-citation achieves a similar effect of pointing to earlier work and logging a citation to it without running the risk of excessive vanity or self-advertising (e.g. Parsons 2015; Shema 2012). It is, moreover, in line with a shift in citations generally from sentential to parenthetical formats, with a preference for the latter increasing by over a third during the last 50 years and now representing 85% of all citations in these four fields (Hyland & Jiang 2016). This, of course, gives the research, rather than the writer, greater prominence, allowing authors to appropriate others' contributions with more cursory acknowledgement. It may also be that a sentential self-citation could risk jarring on readers, drawing more attention than is welcome by the writer. Simply ensuring that previous work is referenced is sufficient to stake a claim for prior work and relevant competence.

There has also been a shift towards less obtrusive self-citation during the period, with a large increase in co-author citations. As noted above, this form of self-citation means that a particular academic may not directly cite him or herself, but is cited by one or more of his/her co-authors (Ioannidis 2015; Glanzel & Thijis 2004). This kind of citation makes little difference to the frequencies in this study as the current authors are self-cited directly and the previous co-author is not relevant to our research. We do not have the sophisticated software needed to identify long-standing research groups who frequently cite each other, but we believe the direct vs co-author distinction is a useful one. It signals that not all authors need to be in the reference for self-citation to occur and helps document a growing trend towards collaborative research in the sciences.

Co-author citation is a consequence of growing academic competitiveness which affects an author's visibility among colleagues and contributes to a citation profile in the indices upon which his or her career might depend. The importance of such co-author self-citation is suggested by the fact that Scopus, the world's largest abstract and citation database of peer-reviewed academic literature, is able to generate citation counts which exclude both direct and coauthor self-citations.

Overall then, self-citations are more numerous than 50 years ago, with frequencies increasing by 46% when allowing for the increased length of papers. As the literature continues to expand at an unprecedented rate, however, they are also less visible today as they comprise a smaller proportion of citations overall. In the next section we turn to look at how authors in different disciplines have contributed to these changes.

5 Disciplinary practices in self-citation

The rise of self-citations during this period displays disciplinary variation. The results show considerable rises in sociology, which has shown a growth of over 400%, from 1.9 cases per 10,000 words in 1965 to 9.8 in 2015, applied linguistics, with a 276% increase (from 2.5 to 9.4) and electrical engineering up 234% (from 5.9 to 19.7). Biology, on the other hand, has seen a decline in self-citations by about 8%, from 20.7 in 1965 to 18.9 in 2015, perhaps explained by the fact that this is the most highly citing discipline of the four and self-citations have not kept pace with the growth in overall citations. Generally, these changes reflect the growing pressures on authors to both establish their credentials by demonstrating proven experience in an area through previous work and by adding to their citations which are harvested by the indexing databases.

Table 4 Changes in self-citation per 10,000 words and as a percentage of citations

	Applied linguistics			**Sociology**			**Elec engineering**			**Biology**		
	1965	**1985**	**2015**	**1965**	**1985**	**2015**	**1965**	**1985**	**2015**	**1965**	**1985**	**2015**
self-citation	2.5	7.9	9.4	1.9	6.2	9.8	5.9	11.8	19.7	20.7	20.9	18.9
All citations	4.4	37.5	124.3	58.6	73.7	152.2	30.5	48.9	102.2	116.9	131.3	170.6
Self/all cites %	56.8	21.1	7.6	3.2	8.4	6.4	19.3	24.1	19.3	17.7	15.9	12.8

The rise may also result from increasing specialization as research becomes ever more expensive, collaborative and focused. Attributing changes in individual disciplines to particular causes is hazardous, not least because of the internal variability of disciplines themselves. But while we hesitate to do so, some factors may be important. In sociology, for instance, there has been to a growing consensus regarding the central problems of sociological theory which have tended to solidify around the dichotomies of subjectivity vs objectivity, structure vs agency, and synchrony vs diachrony (Archer & Tritt 2013). Such a consensus may help unite the literature into general lines of inquiry which facilitate self-citation. The expansion of commercial applications in electrical engineering in the past 50 years helps explain the growth of self-citations in that discipline as researchers specialize in particular subfields such as power, control, electronics, instrumentation and so on. Applied linguistics too has become a much larger and more specialized discipline offering opportunities for researchers to occupy ever smaller niches, so that the Wiley-Blackwell encyclopedia (Chappelle 2012), for example, covers 27 major areas of the field from language learning through bilingualism to policy and planning, each providing clear lines of inquiry to pursue and facilitating self-citation.

We can also see that citations have also shot up overall. This increase is partly due to the ease with which they can now be automatically gathered by bibliometric software as a result of online publishing and technological advances. But, as we noted above, there has also been an exponential rise in the literature itself, with more published research available for citation. The disciplinary variations are also interesting here, with the increases especially marked in the two soft knowledge fields, increasing by 272% in applied linguistics and 160% in sociology. Applied linguistics, of course, has consolidated itself into an established discipline over this period from an emerging sub-field in the 1960s with a limited literature to nearly 600 journals listed on the SCImago rankings[5] today. Sociology has also gained credibility as a discipline and attracted more researchers and journals and, together with applied linguistics, has also seen a shift from book to journal publication, facilitating the easier automatic collection of references to earlier indexed work. Biology, with a 91% increase, has always been a very heavily citing discipline. Adel and Garretson (2006), Samraj (2008) and Hyland (2004) all found that biology was closer to the social sciences in its frequency of citation.

[5] The SCImago journal lists can be found at http://www.scimagojr.com/journalrank.php

Given these changes, the ratio of self- to all-citations becomes interesting. In 1965 self-citations comprised 57% of all citations in our applied linguistics corpus, an astonishing statistic explained by the infant status of the field at that time. Applied linguistics only began to emerge in the late 1960s, slowly establishing its own identity as an interdisciplinary field distinct from formal linguistics and concerned with real-world language problems such as policy, second language acquisition and teaching (Berns & Matsuda 2006). In 1965 there was no distinct literature to speak of and few authors publishing so that self-citation was almost a necessity in gaining traction for these new socially-relevant issues. The subsequent decline in the proportion of self-cites to citations simply reflects the growth of citations overall and brings them more into line with the other disciplines. Biology also saw its ratio fall substantially, so while increasing self-citations, as one might expect from a hard science field where knowledge accumulates by building brick-by-brick, all citations have continued to increase as the field grows and expands.

Table 4 shows that Sociology is the only discipline to display increased frequencies in all three rows between 1965 and 2015, with nearly four times more self-citations, three times more citations and double the ratio between these two over the past 50 years. One reason for this is perhaps that a significant branch of this discipline resembles the hard sciences with an interest in more linear quantitative research. Often termed *instrumental positivism* (Gartell & Gartell 1996), this values methodological replicability, reliability and validity and depends on grounding both theory and methods in a body of existing literature. Obviously this facilitates, even encourages, writers to refer more frequently to their earlier work in the chain of understandings which lead to their present study. Citations have grown together with the growth of the literature in all these fields (except biology), however, and substantial reference lists of 50 items or more are commonplace, providing opportunities for individuals to add to their own linear contributions. Self-citations have risen steadily in electrical engineering but risen and fallen again as a percentage of all citations. Texts in the engineering corpus have the highest proportion of self-citations, with nearly 20% of cites being to one of the current authors.

6 Forms of self-citation

The dependence of arguments in networks of references not only suggests a cumulative, often linear, progression of knowledge, but acts to locate both writers and their claims within a recognised disciplinary framework. How writers choose to represent a connection to their earlier work, then, is of interest as it displays the extent to which this association is foregrounded or left implicit. Writers can, for instance, mention themselves explicitly in the body of the sentence, as in (8), or simply reference their previous work in parenthesis, as in (9):

(8) As in previous work [27], ***we*** considered "sleep" genes to be only those that were upregulated in the sleep group (6–7 hr of sleep) relative to both the spontaneous wake group (6–7 hr at night), and the forced wake group (4 hr of sleep deprivation after spending most of the night awake). [Bio, 2015]

I have attempted earlier (1964) to give a simple description of the form of generative grammars in general and transformational generative grammars in particular as these were described by Chomsky in Syntactic Structures. [App ling, 1965]

In a series of recent publications (Block, 2012, 2014; Block, Gray, & Holborow, 2012), ***I*** have devoted a fair amount of space to defining social class. [App ling, 2015]

(9) To avoid including Twitter users that are not human (companies, institutions, bots, etc), data that have recognisable human characteristics were filtered (as explained in detail in Arnaboldi et al., 2013a and Arnaboldi et al., 2013b) [Soc, 2015]

Some of the trajectories of the equation have been plotted in an earlier paper [10]. [Elec eng, 1965]

Measures of visible light sensitivity in the X-irradiated scotopic receptor confirm the earlier report by Dawson and Smith (10). [Bio, 1965]

Clearly, these decisions may have an impact on how readers process the information they have been given. Sentential self-citations more clearly associate the author with the earlier work by focusing on the actions of the researchers, providing writers with a means of ensuring that their contribution to a

topic is noticed. The parenthetical self-citations, on the other hand, emphasize the research itself, facilitating the automatic collection of citations by indexing software while sidestepping any accent on the fact that a member of the current research team was responsible for that publication. Thus the more cursory acknowledgement of earlier research does not necessarily relegate the cited author to a peripheral space but ensures credit is attributed to the authors, albeit in a more subtle and indirect way.

Table 5 shows a trend in our data towards avoidance of the more self-promotional elements of self-citation in favour of parenthetical forms, with a 100% increase from an average of 7 cases per 10,000 words in 1965 to 14 in 2015, following a similar trend in citations to others' work from sentential to parenthetical (Hyland & Jiang 2016). All disciplines with the exception of biology have hugely increased their use of parenthetical self-citations while sentential forms have held steady or declined. Biology is also the only discipline to see a decline in self-citations over the period.

Table 5 Types of self-citation over time by discipline (per 10,000 words)

	Applied linguistics			**Sociology**			**Elec engineering**			**Biology**		
	1965	**1985**	**2015**	**1965**	**1985**	**2015**	**1965**	**1985**	**2015**	**1965**	**1985**	**2015**
self-citation	2.5	7.9	9.4	1.9	6.2	9.8	5.9	11.8	19.7	20.7	20.9	18.9
sentential	1.7	1.5	0.5	0.1	0.4	0.7	0.8	0.7	0.5	0.6	0.4	0.3
parenthetical	0.8	6.4	8.9	1.8	5.9	9.1	5.1	11.1	19.2	20.1	20.5	18.6
Direct only	2.4	5.6	6.2	0.5	1.7	3.1	3.9	7.3	3.1	12.7	7.7	3.2
Co-author	0.1	2.3	3.2	1.5	4.5	6.7	2.0	4.5	16.5	8.0	13.2	15.7

Establishing a claim to a sustained engagement with a particular research area is also accomplished by authors projecting forward, rather than backwards, in time to promote yet unpublished work:

(10) Bensoussan, Yam, Djehiche, and Tembine (2014), Djehiche, Basar, and Tembine (under preparation) and Djehiche and Tembine (in press) have established a stochastic maximum principle for risk-sensitive mean-field-type control where the key mean-field term is the mean state. [Elec eng, 2015]

The proof will appear elsewhere [8]. [Elec eng, 1965]

It should be noted that there was no evidence for specific localization to either the cytoplasm or the nucleus (A. Catalano & D. H. O'Day, unpublished data; Myre & O'Day, 2002). [Bio, 2015]

The opportunity to alert readers to forthcoming publications is a strategy largely confined to the hard sciences, where research and publication tend to move much faster than the social sciences, and which allow writers to more finely dissect their work, fragmenting single coherent bodies of research into as many publications as possible. This practice, known as "salami slicing" (Abraham 2000), sidesteps strict editorial policies against duplicate publication by breaking a single research paper into their 'least publishable units' (Broad 1981), so that several papers, some in press, may report different findings from the same study.

It is additionally worth observing here that there has also been a change in authors' use of reporting verbs in sentential self-citations with a steady shift towards research acts over the past 50 years. These are verbs which refer to things in the real-world and often occur in statements of findings (*observe, discover, notice, show*) or procedures (e.g. *analyse, calculate, assay, explore*) (Hyland 2004) (see examples in 11). In contrast, discourse acts (12), which involve verbal expression (*ascribe, discuss, hypothesize, state*) have declined by half and cognition acts (13) which are concerned with mental processes (*believe, conceptualize, suspect, view*) by 300%.

(11) *I calculated* the reducibility and schema strength manipulations by reproducing the effects (Brashears, 2013). [Soc, 2015]

We concluded (Burch, Burwell, and Rowell, 1964) from these considerations that the effector for symmetrical mitosis in tissues behind a blood-tissue barrier should generally be found in the a2-globulin fraction. [Bio, 1965]

(12) In (Flowerdew, 2015a) *I mention* that I encourage my postgraduate students to voluntarily express their opinions on the software programs I use and have found that they particularly like the 'high-tech' MICUSP search engine interface. [App ling, 2015]

In my article "A Generative Rhetoric of the Sentence" (CGC, October 1963), *I said* that the principles used there in analyzing the sentence were no less applicable to the paragraph. [App ling, 1965]

(13) In our previous paper (Bourne and Ovinton, 2012) *we predicted* that these changes would occur as a result of overnight temperature fluctuations. [Bio, 2015]

Although frequencies are too small to make any bold claims, with only 110 cases in the entire corpus, ascribing particular research processes to their work, rather than characterising it as an argument or belief, helps strengthen the self-citation by highlighting a proven concrete procedure or outcome. This enables the writer to both demonstrate a productive line of research and promote earlier work which may previously have gone unnoticed.

Finally, the ways authors attribute their self-citations also exhibit disciplinary differences. The shift to co-author types, mentioned above, predominates in all fields except applied linguistics where direct forms comprise almost two thirds of cases. This reflects the greater value attributed to single-authorship in the humanities, where Kagg (2013) notes, for example, 'real philosophers, the type that get tenure, are supposed to fly their intellectual balloons all by themselves'. It also suggests more stable patterns of collaboration, as co-author citations are those where not all the authors of the current paper are authors of the cited paper. There are considerable differences in co-authorship patterns across disciplines. So while experimental sciences such as high-energy physics have an average of nine authors per paper, and often many more, theoretical sciences have far fewer, with physics and computer science averaging three co-authors (Newman 2001). Collaborations of two people are the norm and even this is far less common in the social and human sciences.

Such collaborations help account for the high frequencies of co-author self-citations in the hard sciences, where research groups seem particularly fluid. Inter-disciplinary, and particularly international, collaborations are prized and encouraged, so in1988, for instance, only 8 per cent of articles from the major science and technology regions had international co-authors, but this had risen to 23 per cent by 2009, and later ranging between 27-42 per cent (NSF 2012). Such co-authorship networks form

"small worlds" (Newman 2001) in which pairs of scientists are typically separated by only one or two intermediate acquaintances, a clustering which indicates that scientists introducing their collaborators to one another is an important element of co-authorship and suggests flexibility in writing relationships.

7 Self-citation: practices and changes

Reference to previous work is a prerequisite of academic writing, satisfying the obligation to credit relevant earlier studies, to embed the research and writer in a disciplinary tradition, and to support current claims. Where such reference is to the authors' own prior work citations take on an additional significance, enhancing personal authority and emphasising a sustained engagement in a common professional endeavour. In Fowler and Aksnes (2007) words, "self-citation advertises not only the article in question, but the authors in question" (p. 433). For this reason it is sometimes dismissed as egocentric self-puffery, but self-citations seem to increase with the age and productivity of researchers and institutions continue to include them in staff evaluations.

The diachronic changes we have discussed in this study suggest several trends over the past 50 years, with self-citations rising from an average of 5 to 12 per paper, and an increase from 10 to 15 per 10,000 words of text. This growth, however, is tempered by the huge rise in citations overall, which have doubled in our corpus since 1965, so self-citation has fallen as a proportion of all citations from around 15% to 11%. We also found a shift to parenthetical citations, with only a bracketed or superscript reference to the authors, and to co-author self-citation, where the reference is to a paper which includes one or more of the current authors rather than all of them. Parenthetical citations avoid the authorial prominence which self-mention necessarily entails while ensuring that the previous publication is cited and hopefully noticed, while the co-author self-citation format reflects the growth, and shifting nature, of collaborative research.

Finally, the use of self-citation is subject to disciplinary variation, with the hard knowledge fields in our corpus containing almost twice as many instances as the sociology and applied linguistics papers. The linear progression of research in the hard knowledge fields, where new claims assimilate prior

claims to gradually contribute to the incremental resolution of a research puzzle, means that references tend to be more tightly bound to a given topic. Scientists are more likely to participate in highly discrete and specialised areas of research and this allows writers to draw on their own work to a far greater extent than in soft fields (Becher & Trowler 2001).

In contrast, issues tend to be more diverse and detached from immediately prior developments in the humanities and social sciences, with comparatively fewer opportunities for self-citation. Despite this, Engineering, sociology and applied linguistics have shown the highest rises in self-citations per 10,000 words of text, reflecting the increasing concentration of research into particular key questions and the pressures on academics to carve out a niche for themselves in a focused topic area. In fact, the proportion of self-citations to all citations has actually risen in sociology, where the figure has doubled as areas of the discipline have become more like the hard sciences in the questions they ask and the methods they use to answer them. The figures for applied linguistics are skewed by the explosion of the field in the past 50 years, moving away from language acquisition and teaching to comprise a plethora of issues and publications, with a consequent rise in citations.

Like citations to others, self-citations have increased partly as a consequence of the fact that there is much more research to report now and that electronic publishing, promotional blogging and hyperlinked reference lists make texts so much easier to notice, locate and access. Referencing one's prior work, however, must also be seen in a wider social context, reflecting the influence of two major changes in academic life. These relate, first, to patterns of work which encourage greater specialisation and collaboration in research and publishing and, second, to the growth of an assessment culture which seeks to quantify productivity and 'impact' and distribute academic and professional rewards on this basis. As a rhetorical practice which promotes the individual as well as his or her research findings, self-citation plays an ever more important role in these circumstances, helping researchers to establish their credibility and reputations. We cannot, therefore, see the changes we have discussed here as simply stylistic proclivities or neutral patterns of argument. The growth of self-citation, as a device for the management of self-presentation, is an indexical marker of twenty first century professional academic writing.

References

Abraham, Philip. 2000. Duplicate and salami publications. *Journal of Postgraduate Medicine* 46(2). 67-69.

Ädel Annelie & Garretson Gregory. 2006. Citation practices across the disciplines: The case of proficient student writing. In Perez-Llantada Carmen, Plo-Alastrue, Ramon, Neumann Claus-Peter (eds.), *Academic and professional communication in the 21st century: Genres and rhetoric in the construction of disciplinary knowledge*, 271-280. Zaragoza, Spain: Prensas Universitarias.

Anthony, Laurence. 2011. *AntConc 3.4.3*. http://www.laurenceanthony.net/software.html

Archer, Margaret S. & Tritt, Jonathan Q. 2013. *Rational choice theory: Resisting colonisation.* Oxford: Routledge.

Aksnes, Dag W. 2003. A macro study of self-citation. *Scientometrics* 56(2). 235-246.

Becher, Tony and Trowler, Paul R. 2001. *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines (2nd edn)*. Buckingham: Open University Press/SRHE.

Berns, Margie & Matsuda, Paul K. 2006. Applied linguistics: Overview and history. In Keith Brown (Ed.), *The Encyclopedia of language and linguistics* (2nd edn), 398-401. Elsevier: Amsterdam.

Bonzi, Susan & Snyder, Herbert W. 1991. Motivations for citation: a comparison of self-citation and citation to others. *Scientometrics* 21(2). 245-254.

Bornmann, Lutz & Daniel, Hans-Dieter. 2008. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation* 64(1). 45-80.

Broad, William J. 1981. The publishing game: Getting more for less. *Science* 211(4487). 1137-1139.

Chang, Yu-Ying. 2006. How do well-established scholars cite themselves over their academic careers? Six stories of six well-established scholars. Paper presented at Conference in Honour of John Swales. Ann Arbor: University of Michigan. 22-23 June.

Chappelle, Carol A. 2012. *The Encyclopaedia of Applied Linguistics, 10 Volume Set*. Oxford: Wiley & Blackwell.

Costas, Rodrigo, van Leeuwen, Thed N., & Bordons, María. 2010. Self-citations at the meso and individual levels: effects of different calculation methods. *Scientometrics* 82(3). 517-537.

Cronin, Blaise. 1981. The need for a theory of citation. *Journal of Documentation* 37(1). 16-24.

Cronin, Blaise & Shaw, Debora. 2002. Identity-creators and image-makers: using citation analysis and thick description to put authors in their place. *Scientometrics* 54(1). 31-49.

Fowler, James H. & Aksnes, Dag W. 2007. Does self-citation pay? *Scientometrics* 72(3). 427-437.

Gartell, David C. & Gartell, John W. 1996. Positivism in sociological practice: 1967-1990. *Canadian Review of Sociology* 33(2). 143-158.

Gilbert, Nigel G. and Mulkay, Michael. 1984. *Opening Pandora's Box: A sociological analysis of scientific discourse*. Cambridge: Cambridge University Press.

Glanzel, Wolfgang & Thijs, Bart. 2004. Does co-authorship inflate the share of self-citations? *Scientometrics* 61(3). 395-404.

Glanzel, Wolfgang, Debackere, Koenraad, Thijs, Bart & Schubert, András. 2006. A concise review on the role of author self-citations in information science, bibliometrics and science policy, *Scientometrics* 67(2). 263-277.

Harwood, Nigel. 2009. An interview-based study of the functions of citation in academic writing across two disciplines. *Journal of Pragmatics* 41(3). 497-518.

Harzing, Anne-Wil. 2013. *The Publish or perish book: Your guide to effective and responsible citation analysis*. Melbourne: Tarma.

Hyland, Ken. 2003. Self-citation and self-reference: Credibility and promotion in academic publication. *Journal of American Society for Information Science and Technology* 54(3). 251-259.

Hyland, Ken. 2004. *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor, MI: University of Michigan Press.

Hyland, Ken & Jiang, Feng (Kevin). 2016. Points of reference: changing patterns of academic citation. *Applied Linguistics*.

Ioannidis, John. P. A. 2015. A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. *Journal of Psychosomatic Research 78*(1). 7-11.

Kagg, John. 2013. On writing with others. *The New York Times*. http://opinionator.blogs.nytimes.com/2013/06/24/on-writing-with-others/?ref=opinion&_r=0 (accessed 20 July 2016)

King, Molly M., Bergstrom, Carl T., Correll, Shelley J., Jacquet, Jennifer. & West Jevin D. 2015. Men set their own cites high: Gender and self-citation across fields and over time. Working paper. http://www.eigenfactor.org/gender/self-citation/SelfCitation.pdf (accessed 20 July 2016)

Maliniak, Daniel, Powers, Ryan & Walter, Barbara F. 2013. The gender citation gap in international relations. *International Organization* 67(4). 889-922.

Newman, Mark. 2001. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences of the United States of America* 98(2). 404-409.

NSF. 2012. *Science & engineering indicators.* http://www.nsf.gov/statistics/seind12 (accessed 20 July 2016)

Parsons, Chris. 2015. The dangers of excessive self-citation. *Southern Fried Science*. http://www.southernfriedscience.com/the-dangers-of-excessive-self-citation (accessed 20 July 2016)

Peritz, Bluma C. & Bar-Ilan, Judit. 2001. The inside-outside looking of the field of bibliometrics-scientometrics as reflected in references. In Davis, Mari & Wilson, Concepción S. (eds). *Proceedings of the 8th international conference on scientometrics and informetrics*, 511-522. Sydney: BIRG.

Phelan, Thomas J. 1999. A compendium of issues for citation analysis. *Scientometrics* 45(1). 117-136.

Pichappan, Pit & Sarasvady, Subramaniam. 2002. The other side of the coin: The intricacies of author self-citations. *Scientometrics* 54(2). 285-290.

Public Policy Group. 2011. *Maximizing the impacts of your research: A handbook for social scientists.* LSE Public Policy Group. Consultation draft 3. http://www.lse.ac.uk/government/research/resgroups/LSEPublicPolicy/Docs/LSE_Impact_Handbook_April_2011.pdf (accessed 20 July 2016)

Prichard, Alan. 1969. Statistical bibliography or bibliometrics? *Journal of Documentation* 25(4). 348-349.

Rowe, Martin. 2016. Electrical engineering in the 1960s: The transistor changed everything. *EDN network*. http://www.edn.com/electronics-blogs/rowe-s-and-columns/4442619/Electrical-engineering-in-the-1950s-and-60s--The-transistor-changed-everything (accessed 20 July 2016)

Rousseau, Ronald. 1999. Temporal differences in self-citation rates of scientific journals. *Scientometrics* 44(3). 521-531.

Ruscio, Kenneth P. 1985. Bridging specialisations: Reflections from biology and political science. *Review of Higher Education* 10(1). 29-45.

Safer, Martin A. & Tang, Rong. 2009. The psychology of referencing in psychology journal articles. *Perspectives on Psychological Science* 4(1), 51-53.

Samraj, Betty. 2008. A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes 7*. 55-67.

Schubert, András, Glanzel, Wolfgang & Thijs, Bart. 2006. The weight of author self-citations: A fractional approach to self-citation counting. *Scientometrics* 67(3). 503-514.

Shema, Hadas. 2012. On self-citation. *Scientific American Blog network*. http://blogs.scientificamerican.com/information-culture/on-self-citation (accessed 20 July 2016)

Snyder, Herbert & Bonzi, Susan. 1998. Patterns of self-citation across disciplines (1980-1989). *Journal of Information Science* 24(6). 432-435.

Swales, John. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Thijs, Bart & Glanzel, Wolfgang. 2006. The influence of author self-citations on bibliometric meso-indicators: The case of European universities. *Scientometrics* 66(1). 71-80.

Thomson Reuters. 2012. *Global publishing: Changes in submission trends and the impact on scholarly publishers.* Thomson-Reuters White Paper. http://scholarone.com/media/pdf/GlobalPublishing_WP.pdf. (accessed 20 July 2016)

Ware, Mark & Mabe, Michael. 2012. *The STM report: An overview of scientific and scholarly publishing, (3rd edn)*. Oxford: International Association of Scientific, Technical & Medical Publishers.

White, Howard D. 2001. Authors as citers over time. *Journal of the American Society for Information Science and Technology* 52(2). 87-108.

Appendix 1: Journal list

Applied Linguistics

TESOL Quarterly (1967-)

Language Learning (1948-)

Foreign Language Annals (1967-)

Modern Language Journal (1916-)

College Composition and Communication (1950-)

Sociology

American Journal of Sociology (1895-)

Social problems (1953-)

The British Journal of Sociology (1950-)

American Journal of Economics and Sociology (1941-)

The Sociological Quarterly (1960-)

Biology

The Quarterly Review of Biology (1926-)

Biological Reviews (1923-)

Radiation Research (1954-)

BioScience (1964-)

The Journal of Experimental Biology (1923 -)

Electrical Engineering

Proceedings of the IEEE (1963 -)

Automatica (1963 -)

IEEE Transactions on Automatic Control (1963 -)

IEEE Journal of Solid-State Circuits (1966 -)

IEEE Transactions on Information Theory (1963 -)